DOCUMENT RESUME

ED 189 188 TM 800 370

AUTHOR Gersten, Russell M.; Carnine, Doug
TITLE Measuring Implementation of the Direct Instruction Model in an Urban School District: An Observational Approach.
INSTITUTION Oregon Univ., Eugene.
SPONS AGENCY Office of Education (DHEW), Washington, D.C.
PUB DATE Apr 80
GRANT OEG-00-7507324
NOTE 31p.; Paper presented at the Annual Meeting of the American Educational Research Association (64th, Boston, MA, April 7-11, 1980).

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Achievement Gains; *Classroom Observation Techniques; Compensatory Education; Competency Based Teacher Education; Elementary School Teachers; *Inservice Teacher Education; Primary Education; *Program Implementation; *Skill Development; Teacher Aides; Teacher Evaluation; Teaching Methods; *Test Reliability; *Test Validity
IDENTIFIERS *Direct Instruction Supervision Code; Distar Reading Program; Levels of Use Interview; Project Follow Through

ABSTRACT

During an extensive inservice program, teachers and aides in 20 kindergarten and first grade Follow Through classrooms were observed for mastery of the Direct Instruction Model. Two trained observers used the Direct Instruction Supervisor Code to record behavior in the fall, winter, and spring. Subjects were observed while teaching a 12 minute lesson in reading or oral language. Mastery was hierarchical: most subjects mastered the two basic components (formats and signals) within two months, and the remaining components (corrections, pacing, and student accuracy) by May. Reliability and validity of the Code, and students' achievement gains were also examined. Code scores had high interobserver reliability and low to moderate reliability between fall and winter observations. Scores had moderate concurrent validity with supervisor ratings and discriminated well between high and low implementers. All aides and teachers scored poorly on Loucks' Levels of Use interview for assessing educational innovations; its validity for a performance based model, such as Direct Instruction, was questioned. Correlations with scores on the Comprehensive Tests of Basic Skills were low for pacing and formats, and moderate for the remaining components. Tracing skill mastery over long periods (1-2 years) can offer valuable information for inservice training. (CP)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED189188

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Measuring Implementation of the Direct Instruction Model in an Urban School District: An Observational Approach

Russell M. Gersten
Doug Carnine
University of Oregon

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

R. Gersten

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

Paper presented at AERA, Boston, Mass., April 7, 1980

TM 800370

Acknowledgements

This research was supported in part by U.S. Office of Education Number G00-7507234. Our project officers, Drs. Marianne Scheirer and Jerry Burns offered valuable technical assistance at various stages of the project. Without Susan Green there could not have been a project of this scope. Ralph Green and Phyllis Haddox helped immensely with coordination between the schools and the research staff. Paul Williams and Tom Heiry with careful analyses and reanalyses of the data. Adrienne Allen kept the entire venture together, expertly combining her clinical skills as a teacher trainer with the research perspective of the authors; her knowledge is deeply appreciated.

Abstract

A study was initiated to document and evaluate the process of implementation of the Direct Instruction Model in a large urban school district. The model includes intensive inservice training. A classroom observational code was developed to assess teacher and paraprofessional performance on five variables, and record patterns of skill acquisition for both groups. Acquisition appears to follow a hierarchal system, with most aides and teachers mastering the two basic components of the model within two months and the remaining components within eight months. Implications of the instrument for improved inservice training and understanding the teacher training process are discussed.

In August, 1978, a large urban school district contracted with the University of Oregon Direct Instruction Follow Through Project to supervise the implementation of a highly structured educational model in 20 kindergarten and first grade Follow Through classrooms. The Direct Instruction project represented a very different educational philosophy and system than the one that dominated the previous ten years of Follow Through in that city.

Implementation consisted of an extensive inservice program. Every teacher and aide attended one to two workshops a month where she was taught the principles and practices of Direct Instruction. They were observed in the classroom once a month by program consultants from Oregon Follow Through and about once a week by district resource teachers. They detailed feedback on the teachers' presentation and modeled appropriate techniques. This is a highly directive somewhat controversial approach to teacher training (for details of training see Morimitsu, Note 3).

A study was initiated to explore in detail the process of implementation. Objectives of the study were to: a) empirically describe stages in the implementation process--i.e. highlight trends in acquisition of new skills by teachers, b) explore relationships between teachers' attitudes, motivation and philosophy and classroom practice, c) examine relationships between teacher performance variables specified in the model and student academic growth.

In many ways, the Direct Instruction Model is an ideal candidate for an implementation study. An almost fully articulated system exists

(e.g. Becker, 1977; Becker, Engelmann & Carnine, 1979) specifying in detail what teachers should be doing (e.g. use of signals), saying (highly structured formats), how much time should be spent on reading, arithmetic, oral language each day; and even how fast lessons should be paced. Thus, operationalization of a behavioral coding system was much more direct and simple than with, say, some of the open classroom models analyzed by the Stallings (1973) group in their Follow Through classroom research. Another unique component of the model is that the role of classroom aides is clearly specified--they teach oral language and arithmetic lessons. Thus the same observation system can be used to analyze performance of both teachers and paraprofessionals.

Though several instruments were utilized in the course of the study, the major focus was on the Direct Instruction Supervision Code (Gersten, Allen & Paine, Note 1), a classroom observational instrument. The focus of this paper will be on the development of the code, its reliability (generalizibility) and concurrent validity, and initial findings on patterns of skill acquisition for teachers and paraprofessionals over an eight month period. Often studies of teacher training focus on relatively short intervals, e.g. two to six weeks. We wish to stress the advantages of tracing evolution of teacher/aide performance over a nine month period.

Evolution of the Direct Instruction Supervision Code (DISC)

The code centered on the measurement of five teacher performance variables: Accuracy of Formats, Use of Signals, Use of Correction Procedures, Pacing, and Student Accuracy (see Figure 1 for definitions). These variables are considered essential to true implementation of the

Direct Instruction Model (Becker, Engelmann & Carnine, 1979). Two of the variables are unique to Direct Instruction--accuracy of Distar formats (obviously) and the use of hand signals to insure unison response. The other three variables have broad educational relevance--student accuracy relates to Bloom's (1976) mastery learning paradigm, while use of correction procedures and pacing relate to variables found by Rosenshine and Berliner (1978), Good & Beckerman (in press) and Good, Grouws & Beckerman (in press) to be consistently correlated with high student academic gains in elementary schools. The purpose of the code was to allow for valid, reliable assessment of trends in teacher and aide performance across time. Since three of the five variables have genereralizability well beyond the Direct Instruction model, these results should have broad implications, especially for the field of compensatory education.

A series of studies were conducted over the past 6 years to verify empirically the importance of these performance variables for increased student learning, responding or attending during Distar lessons. Many of these studies utilized single subject methodology (Hersen & Barlow, 1976), usually multiple baseline designs, where one to three teachers were studied in depth over a three week period and functional relationships were determined between aspects of teacher performance and student performance variables (i.e. rate of response, accuracy of response, proportion of students in the group who responded). Three of the implementation variables specified above have been studied individually--the importance of rapid pacing (Carnine, 1976; Carnine & Fink, 1976), correction procedures (Carnine, in press; Siegel, Note 2), use of hand signals (Cowart, Carnine & Becker, 1976), systematic verbal praise (Kryzanowski, 1976), and as a package of

techniques (Carnine, in press b). Maintaining high student accuracy rate has been a target of research by Brophy and Evertson (1976) and others. Accurately following the key ideas in the DISTAR formats has not been studied. The major limitations of these studies is the relatively brief duration (on the average, two to three weeks) and the limited number of teachers involved in the studies.

It thus seemed important to look at these same variables in a natural situation, the Follow Through classroom, for extended periods of time (Snow, 1974) to discover natural trends resulting from inservice supervision, and explore relationships between each variable and student academic gain.

METHOD

Subjects and Setting

Twenty-one teachers and twenty-one aides were involved in the program. All were regular staff members of either a kindergarten (N = 10) or first grade (N = 11) Follow Through classrooms in one of Follow Through's seven schools in a large urban district. All teachers and paraprofessional aides attended monthly inservice workshops and were periodically observed and offered feedback by on-site supervisors (approximately once a week) and program consultants from the University of Oregon (approximately once every three weeks).

Though the potential sample size was 21 teachers and 21 aides, not every teacher or aide was observed in each phase for a variety of reasons. One of the 7 building principals refused to let the teachers in his school be observed, thus lowering the potential sample size to 17. Some teachers and aides were missed in the fall and winter due to illness, or scheduling conflicts. In the spring, the decision was made to intensively observe a representative sample of 6 teachers and 3 aides, with each receiving two to four observations within the same week in May, rather

than attempt to observe the entire sample of 42 for mere one-session each. The rationale for this decision is discussed below.

Procedure

Teachers and aides were observed once in November (N = 14 teachers, 15 aides), and once in either late December or January (N = 13 teachers, 14 aides). Observations in May covered fewer teachers but were more intensive to increase the validity of the observational data (N = 13 teachers, 10 aides).

Teachers/aides were observed in their classrooms while teaching the Distar lessons in reading or language to small groups of 4 to 10 students. The observers were familiar to both the teaching staff and the students.

Each teacher (or aide) was observed for 12 minutes teaching a small group DISTAR lesson in either reading (for the teachers) or oral language (for the aides). Observers were trained DISTAR consultants. Each teaching trial was recorded in one frame with 3 components--teacher presentation --student response--teacher response (see Figure 2). Errors teachers/aides made in presentation (e.g. inaccurate format) or response (failure to correct an error) were noted. Accuracy of student unison responses was also recorded. If even one student failed to respond when a unison response was required, the box was marked NR/No response). At the end of 12 minutes, the teacher's pace (number of learning trials per minute) and accuracy of the other 4 variables (in percentages) could be calculated.

Teachers and aides were observed once in November, once in either December or January, and twice in April or May. To allow for some measure of concurrent validity, teachers were independently given global ratings

on a 1 to 4 scale from two field supervisors. In addition supervisors recorded whether teachers were spending at least 30 minutes per day on oral reading, a measure of allocated time.

RESULTS

Reliability

Interobserver reliabilities on the 5 variables ranged from 4 = .80 to .92 for a random sample of 7 teachers observed by both staff members in the fall and winter. This was deemed adequate.

Preliminary estimates of temporary stability were performed on the fall and winter data. These would tend to be underestimates due to differential teacher growth rates over time. However, the low range of many of these correlations (range of -.07 to .51, median of .25) led us to rethink our design strategy. The extremely low correlations for Format Accuracy and Signal Accuracy were quite likely due to the highly skewed distributions and restricted range of these measures (see Table 1), thus not a cause of serious concern. On the other hand, other classroom studies have also found performance measures for which there is little variability among teachers (Cooley & Leinhardt, 1980), the low reliability and concurrent validity estimates for reinforcement led us to drop this variable from the code. Siegel's (Note 2) earlier work with reinforcement in the Direct Instruction system would tend to support this position; he found no relationship between use of reinforcement and student academic growth.

For the other 3 variables it appeared that the low to moderate correlations of .25, .34, and .51 might be due to uncontrolled sources of variance, such as lesson difficulty, ability level of the children and/or differential teacher/aide growth. This perspective is supported

by Shavelson and Dempsey (1976) review of many teacher observational measures. Our conclusion is that in order to obtain reasonably stable measures of individual differences between teachers, teachers must be observed three to four times in a short time span (e.g. one week) under comparable conditions. This procedure was followed for the spring observations, though it meant observing fewer teaching staff. Though the fall and winter observational data were not reliable enough to afford valid statements about individual differences, by a simple extension of sampling theory, the estimates of mean teacher (or aide) performance should be unbiased and sensitive to improvements in performance of the samples.

Concurrent Validity

Performance scores on each variable during the Winter Observation Period were correlated with Global Implementation Ratings. Each teacher was independently ranked into quartiles representing the degree of model implementation by the two project consultants. The mean ranking was considered his/her global rating (aides were not ranked). Correlations between these scores are each observational score is reported in Table 1. The median correlation was .45. Correlations were low for only two variables--Reinforcement (which was subsequently dropped from the code) and Signals (a variable which may soon be dropped from the code, see below). For the other 4 variables, significant correlations were found, especially for Pacing (.70), Student Accuracy (.68) and Corrections (.51).

Construct Validity: Preliminary Findings

The increased reliability of the Spring Observational Data allowed for some preliminary attempts of construct validity studies. Two procedures were utilized, a) contrasted groups (Anastasi, 1976) in which the scores of high and low implementing teachers were explored, and

b) an analysis of the relationship between implementation indicators derived by DISC scores with those derived by administration of the Levels-of-Use interview (Loucks , Newlove & Hall, 1975).

Contrasted Groups

The three highest-ranked and four lowest-ranked teachers in the same group were contrasted on Spring Observational scores, a somewhat tentative approach to construct validity, but given the limited sample size (and limited budget), the only one feasible (see Table 2). There were suggestive differences, tending to favor the high-ranked teachers, on Pacing (md = 11.9 vs. 9.5), student accuracy (81.5 vs. 70.8), and especially corrections (md = 83.5 vs. 53.3). Formats showed no difference (95.1 vs. 96.85), probably because all teachers had mastered this most basic skill. Signals showed a difference seeming to favor the low ranked teachers (md = 92.1 for the high group vs. 98.9 for the low group). Interestingly enough, this supports some views (e.g. Meyer, 1979) that use of hand signals is a nonessential component of the Direct Instructional Model; the same effect can be achieved by vocal inflection. Applying appropriate correction procedures and maintaining a high rate of student accuracy during the lessons are certainly more sophisticated.

Relationship between Levels-of-Use and DISC

The Levels of Use (Loucks, Newlove & Hall, 1975) interview was administered to all 21 teachers and 21 instructional aides in February by an independent subcontractor, Emrick and Associates. This interview is a relatively widely used instrument for the assessment of educational innovations; it has both clinical and research utility. We wished to

explore the relative pros and cons of using a general instrument like Levels of Use versus model-specific instruments such as DISC.

Levels-of-Use protocols were scored by a trained rater. The results indicated that all 21 instructional aides and 16 of the 21 teachers received a score of 3, indicating mechanical use of the innovation. Mechanical users are teachers still concerned with problems in the mechanics, or technical details of the innovation, i.e., in this case how to follow formats, reinforce correctly, manage the classroom, correct student errors, etc. Five teachers were considered in advanced stages; three received scores of 4A, indicating routine use of the innovation, and two were considered 4B, indicating refined use. Routine users are those who have mastered all the essentials of the innovation, and appear to the interviewer to do so in natural, competent ways. Refined users have mastered the essence of the innovation, and are now adapting to meet the needs of individual students on specific academic projects.

Because of the restricted range of scores, either Pearson or Sprearman correlations seemed an inappropriate means of relating Levels-of-Use to DISC scores. Instead, a "profile" for each of the 5 advanced users was created, examining their spring observational scores on each performance variable, and their supervisor ratings. An unusual pattern emerged; of the five, two had the highest performance scores in the district, and one had one of the very lowest. One of these teachers was never observed because she followed the district's bilingual program of teaching Spanish only to non-English speaking children during the first year. (She used no Distar, thus not implementing the program at all.)

After reviewing the supervisor reports and DISC scores further, the following became clear. Two teachers rated as 4A by Levels-of-Use, were in fact "Routine," highly competent practitioners of the innovation, performing well above mastery level on every DISC variable and receiving the highest supervisor rating. The third "Routine" user scored below mastery on every DISC variable; she rarely used DISTAR, though she was intellectually knowledgable about the principles. Thus, the "Refined" Users included the one teacher who never followed the model, though she was considering so doing (and in fact has implemented it in her classroom in the 1979-80 school year). The other "Refined" user reached mastery level on all but one performance variable, and received a reasonably high supervisor rating.

Several problems with Levels-of-Use should become apparent. The primary one is the lack of variability in the measure; 16 teachers and 21 aides received the same score, though they represented a wide range of levels of implementation. There is also a validity issue; Levels-of-Use interviews seem to detect those with exceptional patterns, but it is possible that knowledgable, highly verbal users may be able to "talk a good line" even when their day-to-day implementation is uneven, or, in some cases, virtually non-existent.

Criterion Related Validity: Some Preliminary Process-Outcome Findings

The ultimate test of the utility of the DISC system and DISC variables is whether teacher performance scores are correlated with the academic performance of students in their class. The validity coefficients reported in Table 3 represent results of preliminary analyses for Grade 1 teachers. Because of the small sample size (only 6 Grade 1 teachers were observed in the spring) and less than ideal reliability, these results must be treated cautiously. Nonetheless, some promising results are apparent.

The predictor variables were the mean spring DISC scores for each performance variable (Format Accuracy, Pacing, Corrections, Student Accuracy during the lesson, Signals), as well as the total DISC score. In addition, the allocation of academic time rating and the general implementation rating were used as predictors.

Dependent variables were (a) post-test scores[1] on all the Reading subtests of the Comprehensive Test of Basic Skills (CTBS)--Letter Sounds, Word Recognition I and II, Comprehension, (b) gain scores on the Wide Range Achievement Test (WRAT) Decoding subtest, and (c) mean scores on the last 2 DISTAR criterion-referenced Reading tests. The unit of analysis was the class mean. Of this group of tests, all seem fairly strongly related to the program goals for Grade 1, although there are some problems with the formats of the two Word Recognition Tests.

Except for the variables Pacing and Formats, moderate correlations are found between the performance variables and the CTBS post-tests and the DISTAR Continuous Progress Test Scores. In most cases, the total DISC score is a better predictor than any one of the performance variables in isolation. With such a small sample size, it does not make sense to exhaustively examine these results. The major problem with Format as a variable is restricted range (Table 1). Pacing of lessons in the spring observations was confounded by a flaw in the design of the observational schedule; some teachers were observed during rapid-paced phonics drills, others during slower-paced story book reading. This has now been rectified.

[1]Children were not pretested on the CTBS, precluding the use of gain scores. (See Cooley & Leinhardt, 1980, for a discussion of the use of gain scores when the class is the unit of analysis.) We considered using residual scores on the CTBS using the WRAT as the predictor, but pre-post correlations were so low and non-significant that there was no sense in using this procedure.

The reasonably high predictive validity for signals has led us to continue to use it in the observational system, even though its concurrent validity measures were weak. The consistently moderate-to-high correlations across all measures for Student Accuracy, Corrections and Total Score are promising. These appear to be the key variables. The Allocated Time measure also shows reasonably high correlations across all measures.

Analysis of Trends

The descriptive statistics derived from the observation code provide a clear profile of skill acquisition for the samples of teachers (N = 13) and aides (N = 14). Both groups follow the same essential pattern, with the paraprofessionals' mean often slightly below that of the teachers for comparable time intervals (see Figure 3 and Table 4).

By November, after two months of inservice training, two skills are mastered by virtually all teachers and aides--use of Distar formats, and use of signals. For Format Accuracy, the teacher's mean is 91.55 percent (median = 99.58). For paraprofessionals, the mean is 99.5 percent and the median is 99.9 percent. This well above mastery criteria, and little progress is shown during the remainder of the year. The same basic pattern emerges for use of signals, except that some aides take longer to master this skill; $\bar{x}$ = 95.44 and $\bar{X}$ = 83.6 in November for teachers and aides respectively. Use of formats and signals seem to be skills that virtually all teachers and most aides acquire within eight weeks.

On the other hand, use of correction procedures, pacing and maintaining student accuracy at a high level appear to be much more subtle, harder to acquire techniques. Acceptable levels determined by DISTAR experts are: 85 percent accuracy for corrections, mean student accuracy of 80 percent; and a pace of 9 trials per minute. November means are well below this

level: $\bar{x}$ = 42.3 percent for corrections, 60.6 percent for student accuracy, and pace of 6.96 trials per minute for the paraprofessionals. November means are similar for teachers. One can chart steady growth through January until May for both groups. By May the mean teacher pace is 10.1 (mastery level), and corrections is close to mastery ($\bar{x}$ = 83.3, md = 33.5). Student accuracy level during Distar lessons is also close to the acceptable level; $\bar{x}$ = 78.1 percent. The paraprofessionals' average scores are slightly lower on all 3 variables--8.36 for pacing, 75.8 for corrections, and 71.2 for student accuracy--indicating a somewhat lower acquisition rate. These three skills definitely appear to require more sophistication, and more supervised training to master.

A striking change is apparent in the use of Allocated Academic Time in the area of reading. In February only 47 percent of the teachers were spending the required at least 30 minutes per day on oral reading (SD = 51). In May, the figure rose to 82 percent (SD = 39). Allocation of time was a key objective in inservice supervision during the winter and spring.

DISCUSSION

It seems clear that a model-specific observational system can be developed with reasonable reliability and concurrent validity to assess the degree of model implementation at the classroom practitioner. This approach has several advantages over a more general instrument such as the Levels-of-Use (LOU) interview (Loucks et. al.); it is performance-based and thus not confounded by the verbal abilities of the teachers interviewed. Furthermore, Levels-of-Use scores provided little variability for the 21 teachers and 21 paraprofessionals interviewed in this study (37 of the 42 received identical scores). There seemed to be several "false positives" in the LOU diagnosis, teachers classified

as high level implementers who, in fact, were nothing of the sort.

The DISC observational system developed so far is in need of refinement; yet even with the moderate-to-low temporal stability estimates (.47 for Total Score, median of .30 for individual performance variables), the validity coefficients are promising. DISC scores correlate moderately well with supervisor ratings (except for scores on Signals), and seem to discriminate well between high and low implementers (especially scores on Corrections, Pacing, and Mean Student Accuracy). There is thus good reason to suspect that with increased observational time per teacher, reliability and hopefully, concurrent validity will increase.

The positive, generally moderate correlations of DISC scores with CTBS post-test scores in Grade 1 offer further support for this position. Prior research (Becker & Engelmann, 1976) has demonstrated functional relationships between each of these variables and student learning over a short time period; the observational system can give information on which variables are strongly correlated with student gain over a full year period.

The descriptive analysis reveals some reasonably clear patterns of skill acquisition across samples of teachers and paraprofessionals. Two sets of teaching techniques (use of Formats and use of Signals) are acquired by virtually all teachers and most paraprofessionals within a two month time span. On the other hand, most teachers seem to take up to eight months to master the following techniques: systematic use of correction procedures, pacing, and maintaining high frequency of correct student responses. Acquisition rate on these variables is even slower for paraprofessionals. There appear to be some strong individual differences in acquisition rate, although this design does not permit fine

grained analysis. However, it is interesting to note that those teachers ranked in the upper quartile by supervisors differed from the lowest quartile only on these three "higher order" variables--Pacing, Student Accuracy, and Use of Corrections. A major goal of the 1979-80 study will be to refine the DISC code so that patterns of individual differences can be explored, and an objective basis for assessment of degree of implementation of the Direct Instruction model be established.

Tracing patterns of teacher development across long periods of time (i.e., 1 to 2 years) can offer valuable information for refinement and streamlining of inservice training. Since several of the variables measured have relevance for models other than Direct Instruction, many of the findings should be generalizable to other teacher training methodologies, especially for models concerned with low-income or special education populations.

Reference Notes

Gersten, R., Allen, A. & Paine, S. Direct Instruction supervision code. Unpublished manuscript. Eugene, OR: University of Oregon, 1980.

Siegel, M. An experimental investigation of teacher behavior and student achievement in the DISTAR instructional system. Unpublished doctoral dissertation, University of Illinois, 1973.

Morimitsu, C. Supervision of Direct Instruction. Follow Through Project, University of Oregon, Eugene, OR, 1979.

References

Anastasi, A. Psychological testing (4th ed.). New York: MacMillan Publishing Co., Inc., 1976.

Becker, W. C. Teaching reading and language to the disadvantaged--What we have learned from field research. Harvard Educational Review, 1977, 47, 518-543.

Becker, W. C., Engelmann, S., & Carnine, D. The Direct Instruction Model. In R. Rhine (Ed.) Encouraging change in America's schools: A decade of experimentation. New York: Academic Press, 1979.

Bloom, B. Human characteristics and school learning. New York: McGraw-Hill, 1976.

Brophy, J. & Evertson, C. Learning from teaching: A developmental perspective. Boston: Allyn and Bacon, 1976.

Carnine, D. W. Correction effects on academic performance during small group instruction. University of Oregon Project Follow Through Technical Report 76-1, Appendix B, 1976b, 9-22.

Carnine, D. W. Effects of two teacher presentation rates on off-task behavior, answering correctly, and participation. Journal of Applied Behavior Analysis, 1976a, 9, 199-206.

Carnine, D. W. & Fink, W. T. Increasing rate of presentation and use of signals in elementary classroom teachers. Journal of Applied Behavior Analysis, 1978, 11, 35-46.

Cooley, W. W. & Leinhardt, G. The instructional dimensions study. Education Evaluation and Policy Analysis, 1980, 2, 7-25.

Cowart, J. B., Carnine, D. W., & Becker, W. C. The effects of signals on child behaviors during DISTAR instruction. University of Oregon Follow Through Project Technical Repprt 76-1, Appendix B, 1976, 55-85.

Good, T. L. & Beckerman, T. M. Time on task: A naturalistic study in sixth grade classrooms. Elementary School Journal, in press.

Good, T. L., Grourvs, D. A., & Beckerman, T. M. Curriculum pacing: Some empirical data in mathematics. Journal of Curriculum Studies, in press.

Herson, M. & Barlow, D. H. Single case experimental designs: Strategies for studying behavior change. New York: Pergamon Press, 1976.

Kryzanowski, J. A. Praise effects of on-task behavior during small group instruction. University of Oregon Project Follow Through Technical Report 76-1, Appendix B, 1976, 29-37.

Loucks, S. F., Newlove, B. W., & Hall, G. E. Measuring levels-of-use of the innovation: A manual for trainers. Austin, TX: Research and Development Center for Teacher Education, University of Texas, 1975.

Meyer, L. Teaching basic skills to low income children: One community's approach. Unpublished manuscript. Eugene, OR: University of Oregon, 1979.

Rosenshine, B. V. & Berliner, D. C. Academic engaged time. British Journal of Teacher Education, 1978, 4, 3-16.

Shavelson, R. & Dempsey, N. Generalizability of measures of teaching behavior. Review of Educational Research, 1976, 553-611.

Snow, R. E. Representative and quasi-representative designs for research on teaching. Review of Educational Research, 1974, 44(3), 265-291.

Stallings, J. A. Follow Through program classroom observation evaluation.. Menlo Park: Stanford Research Institute, 1973.

Figure 1

Definition of Variables
Direct Instruction Supervision Code

1. Format accuracy: The degree to which teachers follow the key words and concepts in DISTAR formats. (Importance: high)
2. Use of signals: The extent to which teachers use signals appropriately to direct students' attention to the learning task. (Importance: low moderate)
3. Corrections: The degree to which teachers immediately correct student errors. (Importance: high)
4. Pacing: The degree to which there is rapid pacing and presentation of lessons and related increases in childrens' attention and/or learning. (Importance: high to moderate)
5. Student Accuracy: For optimal learning, teachers are urged to keep the accuracy of student responses on or about 75%. The purpose of this strategy is to offer children positive, successful learning experiences and to ensure that the entire group is learning, not just the upper half. This concept fits in with mastery learning paradigms advocated by Bloom (1977).(Importance: high)
6. Reinforcement: The appropriate use of social reinforcement is considered crucial to learning. Teachers can err by reinforcing correct responses too infrequently or too frequently. In either case, student motivation and attention to tasks tends to be lower. (Importance: moderate-high)

Figure 2

Sample DISC Observer Form

Stop 11:1?

__teacher/X aide name ANON R (1) A School EMERSON Start 11:0?

supervisor/observer Allen [1] II III Date 05-03-79 Time 14 min

KEY: AP--attention prompted SE--signal error #--task number
RT--repeat task # ____ FE--format error IT-individual turn

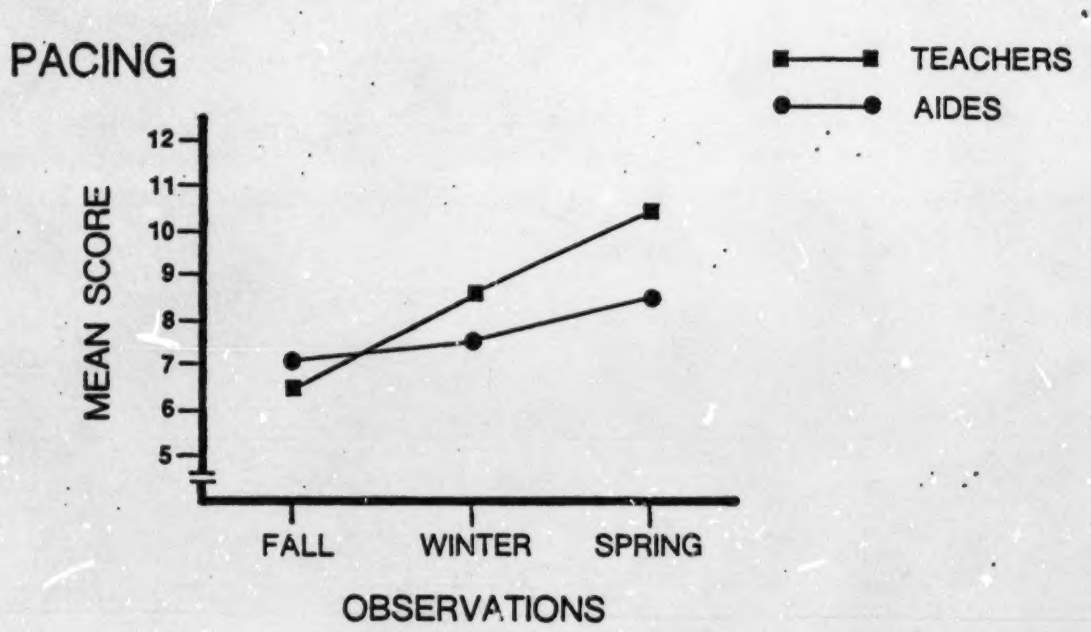
High Group

C--correct student response TO-test-out(corr.resp.on task missed earli
E--error in student response NR-no response by one or more students

GP--general praise EC--error corrected by teacher
SP--specific praise TE--teacher error

AP 01	AP 2	AP 3	AP 4	AP 5	AP 6	AP 7	AP 8	AP 9	AP 1
RT IT	RT IT	RT IT	RT IT	RT IT	RT IT	RT IT	RT IT	RT IT	RT I
SE FE	SE FE	SE FE	SE FE	SE FE	SE FE	SE FE	SE FE	SE FE	SE F
C E	C E	C E	C E	C E	C E	C E	C E	C E	C
TO NR	TO NR	TO NR	TO NR	TO NR	TO NR	TO NR	TO NR	TO NR	TO N
GP EC	GP EC	GP EC	GP EC	GP EC	GP EC	GP EC	GP EC	GP EC	GP E
SP TE	SP TE	SP TE	SP TE	SP TE	SP TE	SP TE	SP TE	SP TE	SP T

Figure 3


PACING
TEACHERS
AIDES
MEAN SCORE
12
11
10
9
8
7
6
5
FALL
WINTER
SPRING
OBSERVATIONS

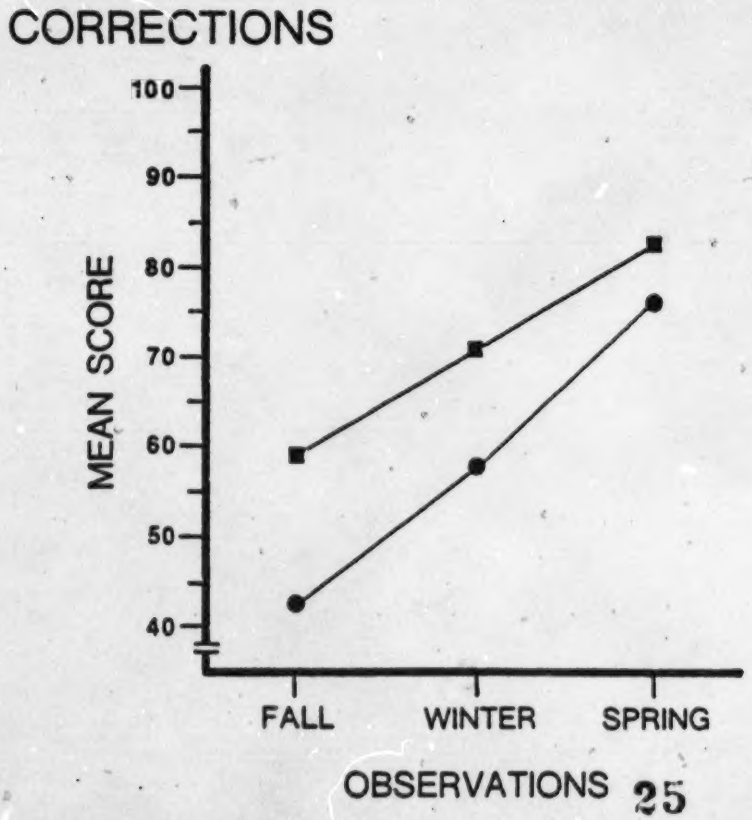
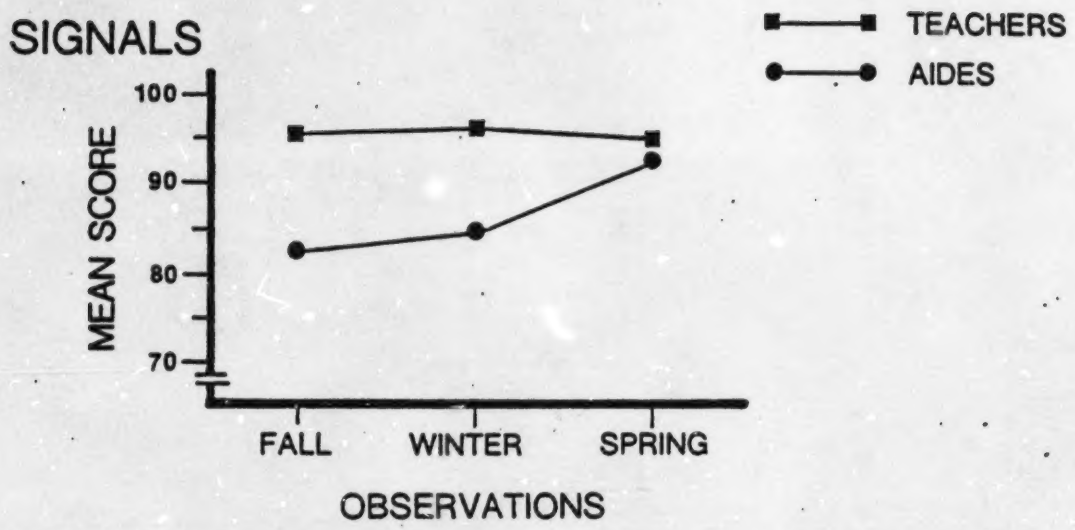
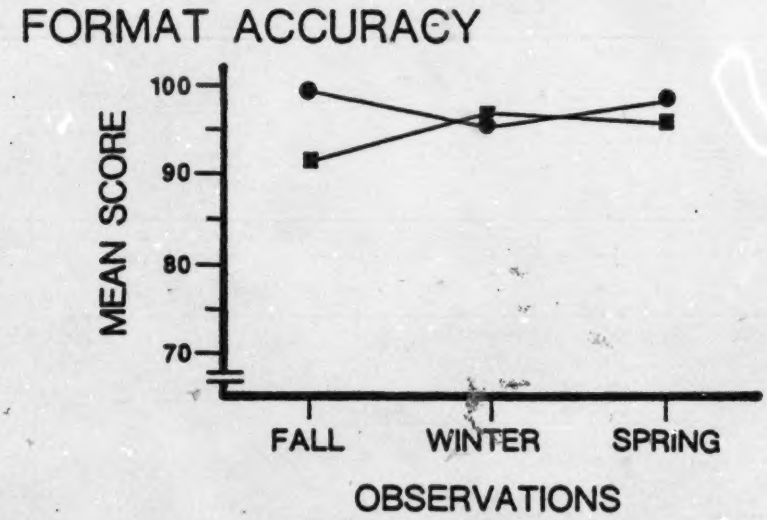

CORRECTIONS
MEAN SCORE
100
90
80
70
60
50
40
FALL
WINTER
SPRING
OBSERVATIONS

Figure 3 (cont'd)





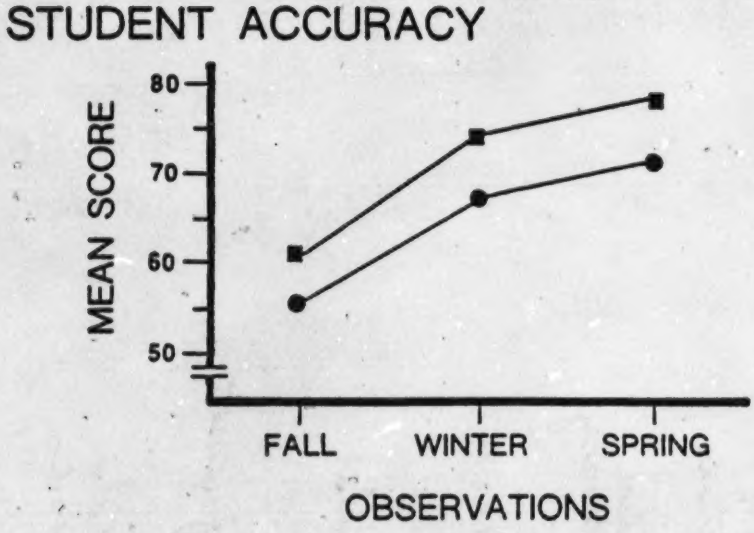


Table 1 Fall-Winter Administration of Direct Instruction Supervision

	Content Validity Rating (1-4 Scale)	First Observation (T 1) (N = 29) November	Second Observation (T 2) (N = 26) Dec. - Jan.	Third Observation (T 3) (N = 23)	Test Retest Reliability (1 week) (N = 23)	Concurrent Validity at T_2	Comments
format accuracy	4	$\bar{X}$ = 95.7 percent SD = 11.4 Md = 99.98 Skew = 3 Kurt = 10	$\bar{X}$ = 95.8 SD = 7.8 Md = -98.95 Skew = -2.8 Kurt = -9	$\bar{X}$ = 97.1 SD = 5.6 Skew = -3.24 Kurt = 12.24	.58*	.37*	Heavy skew. Probably worthless as discriminator between teachers. Interesting descriptive information.
use of corrections	4	$\bar{X}$ = 50.6 SD = 28.7 Md = 49.8 Skew = .13 Kurt = 1.1	$\bar{X}$ = 64.0 SD = 26.6 Md = 64.5 Skew = -.4 Kurt = -.8	$\bar{X}$ = 80.0 SD = 19.1 Skew = -1.90 Kurt = 5.35	.26	.51*	Good variability between SS. A more subtle skills, probably related to high level implementation. Becomes skewed at T.
student accuracy	4	$\bar{X}$ = 58.20 SD = 24.93 Md = 65.3 Skew = -.8 Kurt = -.13	$\bar{X}$ = 70.3 SD = 21.4 Md = 77.1 Skew = -1.4 Kurt = 1.2	$\bar{X}$ = 75.1 SD = 13.1 Skew = -0.78 Kurt = -0.20	.19	.68*	Good variability, approximately normal distribution. Important variable. Significant increase from T 1 to T 2.
pacing	3	$\bar{X}$ = 6.75 SD = 1.73 Md = 6.5 Skew = .35 Kurt = -.2	$\bar{X}$ = 8.09 SD = 1.49 Md = 8.08 Skew = .78 Kurt = .74	$\bar{X}$ = 9.2 SD = 1.47 Skew = -0.41 Kurt = -1.18	.35	.70*	Reasonable distribution and variability. A somewhat less important variable. Increase from T 1 to T 2.
signals	2	$\bar{X}$ = 89.3 SD = 20.1 Md = 97.7 Skew = -2.6 Kurt = 6.67	$\bar{X}$ = 89.7 SD = 17.2 Md = 97.2 Skew = -2.3 Kurt = 5.0	$\bar{X}$ = 94.1 SD = 12.0 Skew = -1.93 Kurt = 2.28	.30	-.22	Skewed distribution; evens out at T3 Not as important as other variables. Consider dropping and delete.
reinforcement	3	$\bar{X}$ = 20.1 SD = 15.4 Md = 18.8 Skew = .98 Kurt = .46	$\bar{X}$ = 15.5 SD = 12.9 Md = 12.7 Skew 1.4 Kurt = 2.8	------	-----	.08	Reinforcement is heavily confounded with format of DISTAR lessons, therefore not a valid performance indicator. DROP
total score		------	------	$\bar{X}$ = 429.2 SD = 42.3 Skew = -1.58 Kurt = 2.24	.47*		

* $p < .05$

* Rate data (all others percent)

Table 2

Contrasted Groups Analysis of Spring Observational Data

	Median Scores for High Implementing Teachers[b] (N = 3)	Median Scores for Low Implementing Teachers (N = 4)
Pacing[a]	11.1 (3)	9.6 (2.9)*
Signals	92.1 (3.5)	98.9 (3.3)**
Format	95.1 (3.5)	96.85 (3)
Student Accuracy	81.5 (2)	70.8 (2)*
Corrections	83.5 (2.5)	53.3 (2)*
Totals	90.4 (3)	71.2 (2.6)*

*Suggestive of difference between groups favoring high implementers.

**Suggestive of difference between groups favoring low implementers

Table 3

Preliminary Process Outcome Correlations, Grade I Teachers (N=6)

DISC	CTBS Post-test Letter Sounds	Comprehension	Word Recognition I	Word Recognition II	WRAT Gain (N=5)	DISTAR Reading Reading - Mean Continuous Progress Test Score
Pacing	.17	-.01	.09	.21	-.21	.51[a]
Signals	.79*	.52[a]	.61[a]	.69[a]	.48	.92*
Formats	-.37	-.35	-.43	-.68	-.31	-.42
Mean Student Accuracy	.69*	.64[a]	.68[a]	.78*	.56	.94*
Corrections	.60[a]	.65[a]	.61[a]	.70[a]	.63[a]	.50[a]
TOTAL DISC	.78*	.62	.66[a]	.74*	.41	.50[a]
Allocated Time	.78*	.69*	.74*	.41[a]	.43[a]	.43[a]
Supervisor Rating	.65*	.50*	.55*	.44[a]	.13	.27

* $p < .05$

[a] Suggestive of positive significant relationship, $.15 < p < .05$

Table 4

Mean and Median Frequencies of Teacher Performance Variables on Direct Instruction Supervision Code

Fall (November, 1979) Observations

	Teachers (N = 14)			Aides (N = 15)		
	Mean	SD	Median	Mean	SD	Median
Pace *	6.52	1.87	6.01	6.96	1.62	6.71
Corrections	59.46	26.05	55.65	42.34	29.6	33.3
Signals	95.44	6.53	98.1	83.6	26.5	97.4
Format Accuracy	91.55	15.5	99.58	99.5	1.55	99.9
Student Accuracy	55.64	24.34	55.45	60.61	26.1	66.7

Winter (December-January Observations

	Teachers (N = 13)			Aides (N = 14)		
	Mean	SD	Median	Mean	SD	Median
Pace *	8.60	1.87	8.86	7.48	1.59	7.67
Signals	95.68	6.49	99.05	84.9	21.8	95.7
Formats	96.36	5.94	99.00	95.55	9.23	98.95
Student Accuracy	74.61	18.87	80.8	66.91	22.96	73.0
Corrections	71.6	18.9	70.3	57.89	30.7	49.2

Spring (May, 1979) DISC Results

	Teachers (N = 13)	Aides (N = 10)
Pace	$\bar{X}$ = 10.3 md = 10.4	$\bar{X}$ = 8.36
Corrections	$\bar{X}$ = 83.3 md = 88.5	$\bar{X}$ = 75.83 md = 80.4
Signals	$\bar{X}$ = 94.6 md = 99.9	$\bar{X}$ = 93.5 md = 99.7
Formats	$\bar{X}$ = 96.0 md = 99.75	$\bar{X}$ = 98.6 md = 99.8
Student Accuracy	$\bar{X}$ = 78.1 md = 77.4	$\bar{X}$ = 71.2 md = 67.7

*Rate data. All others are percentages.